MPEG-4 AVC TRAFFIC ANALYSIS AND BANDWIDTH PREDICTION FOR BROADBAND CABLE NETWORKS

A Thesis
Presented to
The Academic Faculty

by

Laetitia I. Lanfranchi

In Partial Fulfillment
of the Requirements for the Degree
of Master with Thesis Option in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2008

MPEG-4 AVC TRAFFIC ANALYSIS AND BANDWIDTH PREDICTION FOR BROADBAND CABLE NETWORKS

Approved by:

Benny K. Bing, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Fred B-H. Juang, Co-Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Gee-Kung Chang
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: June 25^{th} 2008

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Video transmission requires a large amount of bandwidth resources. Thus, bandwidth allocation and bandwidth prediction are significant challenges that merit investigative research. In this thesis, we study the bandwidth requirement of the MPEG-4 AVC standards and propose new algorithms for bandwidth prediction.

After an introduction of the subject in Chapter 1, we analyze MPEG-4 AVC compressed high definition (HD) video traffic. The construction of a group of pictures (GOP) and the importance of each type of frame are described and explained in Chapter 3. We study the frame size variability of the different frame types. The short term autocorrelation and crosscorrelation, and the long range dependence characteristics of the traffic are then quantified.

The analysis of traffic leads to the design of several prediction algorithms. We propose a model to predict the B-frame and the GOP size in the short term in Chapter 4. The models are designed to be simple (in order to be implemented in a real-time fashion) but accurate. The models fit the traffic well, and at the same time, capture the marginal error distribution effectively. These models can be enhanced using a new scene change detector. A model for the error is also proposed. The error model is used as a performance metric to evaluate and compare the various models proposed.

In Chapter 5, we propose an algorithm that can be used to predict, in the long term, the size of the I-frames, the P-frames, the B-frames, and the GOP. This model is more complex than the previous ones, but this drawback is offset by the additional time available for long-term prediction. A discussion on the appropriate parameters is provided. We analyze the results and compare them to the ones obtained with short-term prediction.

In Chapter 6, we apply the algorithm presented in Chapter 5 to MPEG-2 video. We also study the influence of the quantization parameter on the results. The quantization parameter determines the quality of the video.

The last chapter (Chapter 7) briefly presents network coding and its relation to video frame transmission. Network coding can be used to reduce the bandwidth requirements, and improve the robustness of video transmission.

CHAPTER 1

INTRODUCTION

1.1 Broadband Cable Networks

A popular installation for broadband cable networks is the Switched Digital Video (SDV) architecture. In this architecture the server sends video signals only to set-top boxes (STBs) that tune into them, thereby conserving the aggregate network bandwidth by not broadcasting signals to all STBs all the time (Figure 1). In addition, significant cost benefits can be achieved from bandwidth sharing and optimization through high density video processing. SDV allows each node or region to operate with a level of programming complexity once only reserved for main distribution centers. It moves complex processing of the channel lineup closer to the subscriber, placing heavy demands on the edge video processing equipment.

Another new installation is the Video-On-Demand (VOD) architecture. While SDV typically deals with real-time video, a video-on-demand (VoD) service focuses on pre-recorded video. VoD subscribers can view selected movies or TV programs on demand. VoD traffic is currently small compared to broadcast TV but may use substantial amounts of bandwidth since VoD streams are normally unicast.

The video traffic in SDV and VOD systems primarily consists of large MPEG file transfers, which can be classified regarding the definition level of the video. Such traffic is characterized by high peak rates and drastic short-term rate changes. Because of these drastic changes, the design of a STB's buffer well adapted to the traffic is challenging. The STBs' buffers sizes are either too large - and not well utilized - or can easily be saturated, which can lead to undesirable losses. With the SDV and VoD architectures, the bandwidth demands are asymmetric. The so-called digital video services normally occupy more downstream than upstream bandwidth. For the need to support packetized video, the video traffic can be set real-time constant bit rate (CBR) or variable bit rate (VBR) with peak viewing hours.

Several video standards exist. Moving Picture Experts Group 2 (MPEG-2) is the most prevalent video compression standard in SDV and VoD installations today. However the real compression savings of MPEG-4 AVC are obtained from compressing high definition video. Since HD video will grow in importance in the future, we will see a corresponding demand for MPEG-4 AVC compressed

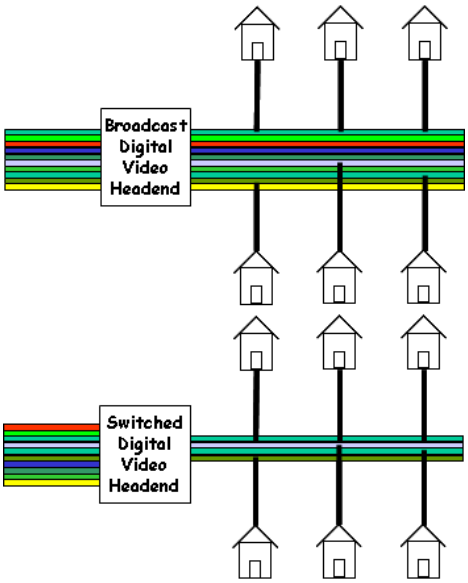


Figure 1: Switched digital video architecture

video.

Both SDV and VoD cable service architectures must currently receive CBR video streams to operate seamlessly. The CBR traffic simplifies and maximizes the utilization of each available QAM signal at the physical layer. Many operators tend to use a set of preset bit rates for various streams based on the video complexity of those streams. The main drawback for doing this is that it does not take advantage of new compression standards such as the MPEG-4 standard which has high bit rate variability. Thus, if the fixed bit rate is set too low to support the MPEG-4 standards, it can lead to packet losses that impair the video quality dramatically. Conversely, if the fixed bit rate is set too high, then the channel bandwidth is underutilized. This is why a dynamic scheme would be more suitable for bandwidth allocation. A dynamic scheme for bandwidth allocation must match on the characteristics of the transmitted video.

1.2 Management and Bandwidth prediction

One key challenge facing many broadband cable network (BCN) operators is the need to optimize bandwidth resources for video transmission without incurring packet losses that compromise video quality playback at the set top box (STB). Bandwidth is actually the most expensive resource in video transmission. The prediction of the trace of a video can be used to monitor and manage bandwidth utilization at the cable headend. This allows a cable headend to provision bandwidth efficiently or reduce STB buffer requirements and packet losses under periods of peak network utilization. As will be explained in Chapter 3, bandwidth prediction can also be use for selective B-frame dropping and

bandwidth conservation.

CHAPTER 2

PREVIOUS WORK

The traffic chracteristics of compressed video has attracted a lot of interest during the past decade. The network transport of compressed video is impacted by these characteristics. In [1], the authors study the characteristics of MPEG-4 AVC video traffic. They examine the bit rate distortion performance, bit rate variability, and long range dependence of the MPEG-4 AVC codec for videos with various resolutions. For the purpose of provisioning bandwidth, the traffic needs to be modeled. In [2], the authors use a periodicity transform to identify the most significant periods of the traffic and use an autoregressive time series to capture the autocorrelation and apply the G-and-H distribution to model the marginal distribution. In [3], the authors analyze the autocorrelation function, the Gamma-shaped probability density function and the correlation between different frames belonging to the same GOP. This analysis leads to a Markov-based model for MPEG video traffic. The model contains two levels which capture both the inter-GOP and the intra-GOP correlation.

For efficient bandwidth allocation, many dynamic schemes have been proposed in the literature. In [4], the authors develop an efficient video assignment mechanism for maximizing the profit of a VoD system. In addition, prior work on short-term linear prediction of MPEG video traffic has been reported. In [5], the authors propose a scheme which predicts the composite MPEG video traffic levels. The scheme is based on predicting the relative change in the frame sizes between adjacent GOPs. In [6], the authors propose a model which partitions the vector of MPEG frames according to scene changes to improve the accuracy of the prediction. In [7], the impact of long-term dependence is shown, in terms of its implications for network design and network control strategies.

However, most of these predictive schemes were designed based on the traffic characteristics of pre-recorded videos encoded using the older MPEG-2 or MPEG-1 standards. In this thesis we focus on the MPEG-4 AVC standards.

CHAPTER 3

MPEG-4 AVC TRAFFIC ANALYSIS

The objective of the traffic analysis is to find an appropriate model for the bandwidth prediction. In section 3.1, we present the different frames of a MPEG file and the GOP structure we use. Then we compare some statistical characteristics of the different frames, the GOP and the whole video in section 3.2. In section 3.3, we provide insights into B-frame dropping, which is the direct consequence of the statistical analysis. Finally, we focus on the correlation of the frame sizes in the short-term in section 3.4 and on the long range dependence characteristic in section 3.5.

3.1 GOP Structure

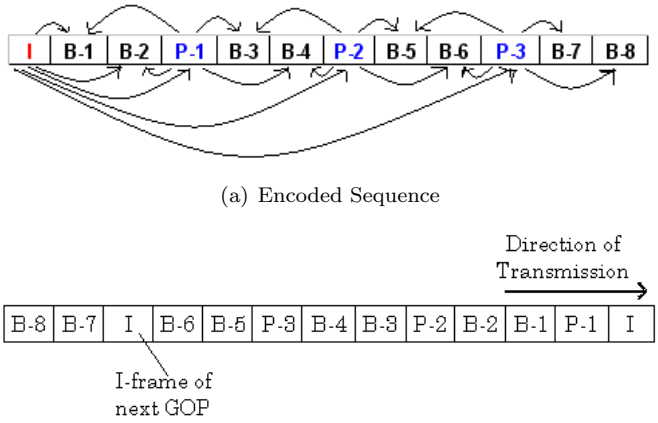
When a video is compressed using MPEG standards, the encoded video becomes a succession of three frame types: I- P- and B-frames. These frames are aggregated as a Group of Picture (GOP). In this thesis we have focused on a specific structure, the G12/B2 structure.

G12/B2 comprises 12 frames with 2 B-frames in a row between each reference frame. The reference frames are the I- and P-frames. A GOP contains only one I-frame. Thus, G12/B2 includes 1 I-frame, 3 P-frames, and 8 B-frames (see Figure 2(a)). We introduce the following notations. B-1 is the first B-frame of a GOP, B-2 the second one and so on until B-8. We use the same notation for the P-frames: P-1, P-2, and P-3. The I-frame is denoted as I.

The I-, P- and B-frames differ from the way they are encoded. The I-frames are encoded as single pictures, with no temporal information whereas the P- and B-frames contain some temporal information. The P-frames are encoded based on the I-frame. The I-frame is a reference frame of each P-frame of a GOP. Thus, the P-frames contain some temporal information. Each B-frame has two reference frames: either an I-frame and a P-frame (for B-1 and B-2) or two P-frames (for the others B-frames). They contain only temporal information, including some predictive information. In Figure 2(a) the arrows indicate the frame references.

We also define the GOP size as the sum of all the frame sizes in a GOP. Hence, for a G12/B2 GOP structure, the GOP size is the sum of the size of the I-frame, the 3 P-frames, and the 8 B-frames.

The transmitted sequence is displayed on Figure 2(b). The frames are not transmitted in the same order as they are shown when a movie is played. This is due to the fact that the decoder needs, for example, the P-1 frame to decode the B-1 and B-2 frames even if the P-1 frame is played



(a) Encoded Sequence

(b) Transmitted Sequence

Figure 2: GOP structure with G12/B2.

back after the B-1 and B-2 frames. So the P-1 frame is transmitted before the B-1 and the B-2 frames. The same happens for the other B-frames and their P-frame references. Note that for the final GOP, frames B-7 and B-8 are not transmitted. Further details on the construction of a GOP are described in [3].

The different kinds of frame do not have the same importance in a video stream. A B-frame contains less information than a P-frame, which contains less information than an I-frame. Moreover, as we shall see in section 3.2, the different frame types differ from each other by their statistical characteristics. As a consequence, in this thesis, we treat them separately before comparing the results. However, their statistical characteristics are not independent. Some strong correlation between the different frames exists. These properties are exploited in Chapter 4 and 5.

3.2 Frame Sizes and Coefficient of Variability

Although MPEG-4 AVC is more efficient than MPEG-2 (MPEG-4 AVC leads to higher compression efficiency and hence, higher bandwidth savings), an MPEG-4 AVC compressed video stream presents higher bit rate variability compared to MPEG-2 [1]. This rate variability can sometimes be represented by the coefficient of variability (CoV) of the frame sizes, which is computed as follows. For a video sequence consisting of M frames encoded with a given quantization level, if X_t $(t = 1, 2, ..., M)$ denotes the sizes of the encoded video frames, and M is the length of the video in terms of frames, then the CoV of the encoded video is defined as equation 1. In this case, σ is the standard deviation

and $\bar{X}$ is the mean of the frame sizes.

$$CoV = \frac{\sigma}{\bar{X}} = \frac{\sqrt{\frac{1}{M-1} * \sum_{t=1}^{M} (X_t - \frac{1}{M-1} \sum_{k=1}^{M} X_k)^2}}{\frac{1}{M-1} * \sum_{t=1}^{M} X_t} \quad (1)$$

We have computed the CoV of different frame types of several MPEG-4 AVC video traces obtained from [8]. The results are shown in Table 1(a). Videos encoded with a high quantization parameter (QP) value, which corresponds to low quality (LQ) resolution. These videos exhibit a higher CoV than videos encoded with a low QP value or equivalently high quality (HQ) resolution. This can be explained from equation 1. For LQ videos, the frames sizes are smaller, which result in a smaller mean and as a result, the CoV tends to be higher. A higher CoV, complicates the prediction. As a consequence, the prediction is less accurate for the LQ movies that the HQ movies.

From Table 1(a), it is clear that the size of the I-frames are much bigger than the P-frames which are much bigger than the B-frames. The size of a frame reflects the amount of information they contain. As a consequence the loss or drop of a B-frame has less impact on the video quality than the loss of a P- or I-frame.

From Table 1(a), it can be observed that for all MPEG-4 AVC videos, the CoV for the B-frames is higher than the I- and P- frames while the CoV for the P-frames is higher than the I-frames. From Table 1(b), it is clear that if the B-frames are removed from the entire encoded video, the CoV is reduced. Intuitively, this is expected since the remaining I- and P-frames lead to less variation in the frame size. The CoV reduction is consistent across all MPEG-4 AVC encoded videos and is equivalent to smoothing the encoded video bit stream. For the case of the whole video without the B-frames, the mean and standard deviation are computed using 4 frames (1 I- and 3 P-frames) and not 12 frames. Moreover, as can be seen from the standard deviation and the mean, HQ videos demand the highest bandwidth and generate the widest range of bit rate requirements. This is as expected and is consistent with Figure 4 of reference [9].

To summarize, for HQ movies, the standard deviation is an important metric that determines the frame size variation and hence, the range of bit rate requirements. For such movies, in addition to the high average bit rate requirement, the range of bit rate variation is also high. Conversely, for LQ movies that typically require low average bit rates, the CoV is a good metric that describes the bit rate variation. In this case, a high CoV may not necessarily result in a wider range of bit rate variation when compared to high-action or HQ movies.

Since bandwidth allocation for videos with high bit rate variability is very challenging, our solution is to transmit HQ videos (with small CoV) whenever possible but reduce the bandwidth requirements and smooth the encoded video bit stream using selective B-frame dropping. Here, the

Table 1: Statistics for movies encoded with MPEG-4 AVC with G12/B2.

(a) I-, P-, and B-frames

Quality	B-frames				P-frames				I-frames		
	Std dev	Mean	Cov	Mean per GOP	Std dev	Mean	Cov	Mean per GOP	Std dev	Mean	Cov
						Terminator 2					
QP=10	2.8E+05	7.7E+05	0.3698	6.1E+06	3.7E+05	1.3E+06	0.2913	3.8E+06	4.2E+05	1.5E+06	0.2710
QP=28	3.5E+04	3.9E+04	0.9159	3.10E+05	6.8E+04	1.2E+05	0.5602	3.66E+05	9.2E+04	2.1E+05	0.4374
QP=48	2.8E+03	2.3E+03	1.1996	1.9E+04	9.9E+03	1.5E+04	0.6428	4.6E+04	1.7E+04	3.6E+04	0.4636
						Sony					
QP=48	1.0E+03	9.0E+02	1.1518	7.2E+03	8.7E+03	9.1E+03	0.9579	2.7E+04	3.0E+04	5.7E+04	0.5261
						From Mars to China					
QP=28	5.3E+04	5.6E+04	0.9365	4.5E+05	1.7E+05	2.5E+05	0.6641	7.5E+05	3.8E+05	7.4E+05	0.5194
						Horizon Talk show					
QP=28	9.9E+03	1.1E+04	0.8872	9.0E+04	3.3E+04	6.3E+04	0.5239	1.9E+05	9.4E+04	3.4E+05	0.2796

(b) Whole video and GOP

Quality	Whole Video			Without B-frames			GOP (G12/B2)		
	Std dev	Mean	Cov	Std dev	Mean	Cov	Std dev	Mean	Cov
					Terminator 2				
QP=10	4.2E+05	9.5E+05	0.4409	4.0E+05	1.3E+06	0.2991	3.5E+06	1.1E+07	0.3074
QP=28	7.5E+04	7.4E+04	1.0185	8.4E+04	1.4E+05	0.5833	4.8E+05	8.9E+05	0.5442
QP=48	1.2E+04	8.4E+03	1.4746	1.5E+04	2.1E+04	0.7282	5.4E+04	1.0E+05	0.5372
					Sony				
QP=48	1.8E+04	7.6E+03	2.3765	2.7E+04	2.1E+04	1.2658	4.5E+04	9.1E+04	0.4896
					From Mars to China				
QP=28	2.4E+05	1.6E+05	1.4929	3.2E+05	3.7E+05	0.8594	1.1E+06	1.9E+06	0.5722
					Horizon Talk show				
QP=28	9.4E+04	5.1E+04	1.8431	1.3E+05	1.3E+05	0.9914	2.1E+05	6.1E+05	0.3429

benefits are three-fold: better quality video, and reduced bandwidth requirements and CoV. It will be shown in Chapters 4 and 5 that our proposed prediction models are more accurate for HQ than LQ videos. This provides another incentive for sending HQ videos.

3.3 B-frame Dropping

The B-frames occupy the most number of frames (and typically, the most bandwidth) in an MPEG-4 AVC GOP. In addition, the trace of the B-frames tends to be burstier than the trace of the other frames, which implies the bandwidth allocation for the B-frames is more challenging than the other frame types. On the other hand, B-frames are the least important compared to the I- and P-frames. Dropping B frames is less harmful because the frames transmitted following a B-frame are not dependent on that B frame (Figure 2). In addition, B-frames only contain temporal information and so their loss only causes motion artifacts which may be difficult to notice unless the loss rates are very high. On the other hand, random frame loss can cause artifacts randomly in both temporal and spatial domains, and are more observable at lower loss rates. Therefore, B-frames can sometimes be dropped to conserve bandwidth resources and reduce bit rate variability. Reducing the bit rate variability is equivalent to smoothing the encoded video bit stream. In some instances, all or a large portion of the B-frames in the entire video can be removed without compromising the quality of video playback [10].

To further determine the amount of possible bandwidth savings associated with B-frame dropping, we compare the sizes of the B-frames against the size of the I-frames. We observe, from Table 1(a), that the mean size per GOP of the B-frames in a video can exceed the mean size of the I-frames. This implies that the total size of all B-frames in a video can exceed the total size of all I-frames. This is especially true for HQ movies. As expected, the average size of the B-frame is less than the average size of the I-frame for all videos. These observations indicate that substantial bandwidth savings can be achieved in the removing of B-frames, with Table 1(b) confirming our observations.

Thus, the ability to predict the size of the B-frames can either allow a cable headend to provision bandwidth efficiently or reduce STB buffer requirements and packet losses under periods of peak network utilization.

For more detailed information about B-frame dropping, see [11].

3.4 Short term Correlation

In this section, we point out the correlation between the frames of consecutive GOPs. Our short term prediction is based on the correlation.

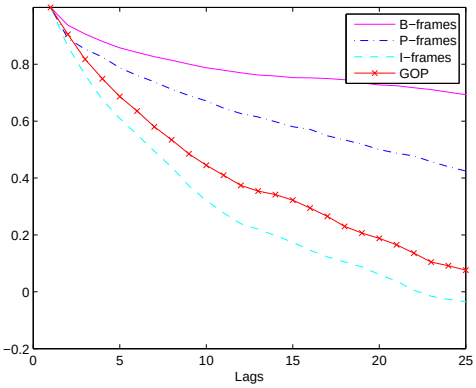


Figure 3: Autocorrelation for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=28 (25 frame-lags).

3.4.1 Autocorrelation of the Frames in the short term

In order to evaluate the autocorrelation of the frames, we compute the autocorrelation r_X of the trace of each frames using equation 2. Here, X represents either the I-frames, P-frames, B-frames or GOP size, σ_X represents the standard deviation of X, $E(.)$ is the expectation and k is the number of frame-lags.

$$r_X(k) = \frac{E[(X_t - \bar{X})(X_{t+k} - \bar{X})]}{{\sigma_X}^2} \tag{2}$$

The values of the autocorrelation of the I-, P-, B-frames and the GOP for the movie Terminator 2 are very close to 1 for the first lags (Figure 3). This shows a strong correlation in the short term. This correlation is very natural because the frames are very close to each others. They mostly belong to the same GOP, and since we observe the correlation for each frame type separately, they are encoded in the same way. It is obvious that this correlation can be exploited to predict the size of the next frame.

3.4.2 Crosscorrelation of the Frames within a GOP

Even if the autocorrelation is very strong, we wish to study other correlations in the same GOP. These correlations have an huge impact on the B-frame and GOP size prediction. Because the B-frames are of particular interest (due to the possibility of B-frames dropping), we first focus on these frames. Then, we observe the crosscorrelation involving the GOP sizes.

As explained earlier, the B-frames are constructed based on the I- and P-frames. As a consequence, the size of the B-frames may be strongly correlated with the size of the reference frames. This short-term correlation is obvious by observing the first 25 lags (in frames) of the crosscorrelation between the B-frames and some of their reference frames as illustrated in Figure 4.

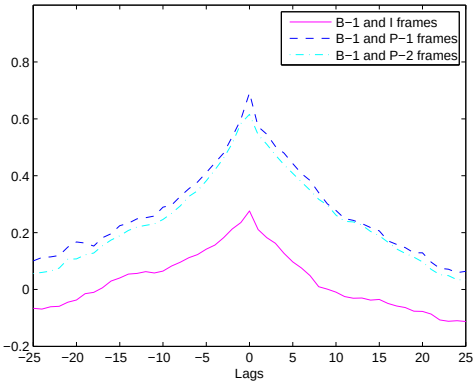


Figure 4: Crosscorrelation for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=28 (25 frame-lags).

Table 2: Correlation coefficients of different frames of videos encoded with MPEG-4 AVC with G12/B2.

	B-1	**B-2**	**B-3**	**B-4**	**B-5**	**B-6**	**B-7**	**B-8**
				Terminator 2 QP=10				
I	0.706	0.709	0.691	0.691	0.659	0.665	0.644	0.644
P-1	0.859	0.882	0.861	0.853	0.822	0.822	0.798	0.796
P-2	0.810	0.824	0.854	0.871	0.862	0.854	0.824	0.819
P-3	0.780	0.796	0.818	0.835	0.851	0.866	0.844	0.846
				Terminator 2 QP=28				
I	0.276	0.290	0.276	0.271	0.256	0.271	0.269	0.265
P-1	0.691	0.707	0.684	0.671	0.646	0.650	0.634	0.626
P-2	0.616	0.627	0.655	0.662	0.645	0.645	0.625	0.615
P-3	0.583	0.598	0.621	0.631	0.654	0.671	0.656	0.646
				Terminator 2 QP=48				
I	0.249	0.254	0.285	0.242	0.289	0.283	0.275	0.244
P-1	0.661	0.687	0.659	0.631	0.641	0.636	0.593	0.553
P-2	0.587	0.593	0.690	0.636	0.662	0.659	0.609	0.575
P-3	0.545	0.557	0.613	0.592	0.688	0.671	0.654	0.614
				Sony QP=48				
I	0.149	0.128	0.041	0.050	0.126	0.130	0.089	0.109
P-1	0.542	0.567	0.602	0.601	0.540	0.515	0.563	0.538
P-2	0.503	0.466	0.721	0.702	0.574	0.549	0.609	0.576
P-3	0.563	0.517	0.648	0.649	0.616	0.612	0.612	0.589
				From Mars to China QP=28				
I	0.394	0.4005	0.397	0.3943	0.3913	0.3886	0.39	0.377
P-1	0.8392	0.8457	0.8232	0.8091	0.7729	0.7679	0.76	0.749
P-2	0.8068	0.8159	0.8402	0.8405	0.8097	0.8037	0.79	0.783
P-3	0.7727	0.7763	0.8011	0.7992	0.8144	0.8218	0.82	0.81
				Horizon Talk show				
I	0.1511	0.1518	0.1565	0.1681	0.1441	0.1759	0.19	0.188
P-1	0.8007	0.8276	0.7177	0.7356	0.6731	0.6734	0.61	0.612
P-2	0.7075	0.7104	0.7872	0.7882	0.7122	0.7031	0.63	0.628
P-3	0.7023	0.6982	0.7394	0.7697	0.7876	0.8301	0.74	0.737

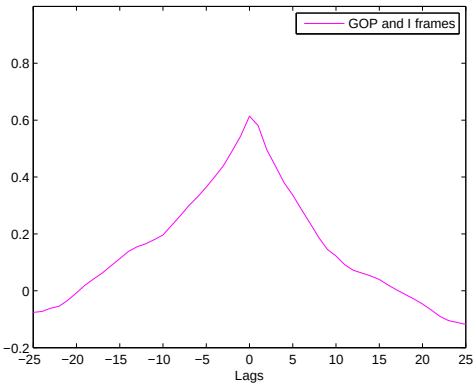


Figure 5: Crosscorrelation of somes frames of a GOP and the GOP for Terminator 2 encoded with MPEG-4 AVC G12/B2 QP=10 (25 frame-lags).

In order to locate the most relevant correlation in the short term, we compute the coefficient of correlation ($\rho_{X,Y}$) between each B-frame (denoted as variable X), and each I- or P- frame (denoted as variable Y) in a GOP using equation 3. Here, $\hat{X}$ and $\hat{Y}$ represent the mean size of the frames while σ_X and σ_Y represent the standard deviation of X and Y respectively.

$$\rho_{X,Y} = \frac{E(XY) - \bar{X}\bar{Y}}{\sigma_X \sigma_Y} \tag{3}$$

The results for several videos are presented in Table 2. The coefficients show a strong correlation between a B-frame and the closest P-frame. The correlation between the B- and I-frames is weaker. This is also confirmed by the plots in Figure 4. Even if some of the B-frames (the first two B-frames of each GOP) are encoded with an I-frame as reference, the coefficient of correlation between the B-1/B-2 frames and the I-frame is lower than the coefficient of correlation between the B-1/B-2 frames and the P-1 frames. This is due to the way the frames are encoded. For instance, the P-frame contains predictive information (that can be used to predict the B-frames) whereas the I-frame is encoded as an image. This leads us to consider the size of the P-frame when predicting the size of the B-frames.

We now study the crosscorrelation for the GOP. The GOP is the sum of all the frames contained in a GOP. According to the statistics presented in Tables 1(a) and 1(b), the biggest frame in a GOP is the I-frame, which is also the first frame transmitted in a GOP. So, the crosscorrelation between the I-frame size and the GOP size may be more important than the crosscorrelation of the GOP with any other frames in the GOP. This is shown in Figure 5. As we can see, the correlation between the GOP and the I-frame is strong.

This leads us to consider an I-frame to predict the size of the GOP. However, because we need

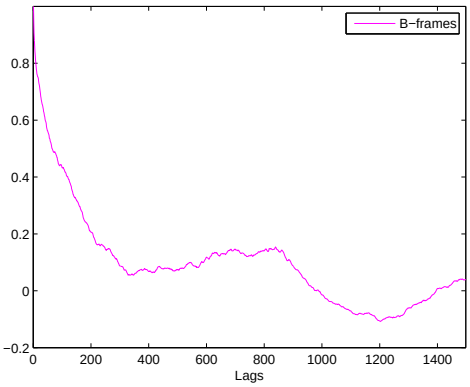


Figure 6: Autocorrelation for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=28 (1500 frame-lags).

to receive a frame before being able to use it as a predictor, we cannot use the size of the current I-frame to predict the size of the current GOP. We have to consider the size of the previous I-frame to predict the size of the current GOP.

3.5 Long-Term Correlation and Long Range Dependence

The long-term correlation and the long range dependence are used for the long-term prediction. In this case, the notion of GOP loses its importance since many scene changes may occur. As a consequence, we do not study the crosscorrelation in the long-term.

3.5.1 Autocorrelation of the Frames in the Long-Term

The long-term correlation is shown by the plot of the autocorrelation of the frames for lags up to 1000 frame-lags. On Figure 6, you can observe that the autocorrelation decrease less exponentially. This shows a long-range dependence characteristic that can be exploited.

3.5.2 R/S Analysis and Hurst Parameter

Another way to observe the long range dependence is to compute the Hurst parameter of the video. The Hurst parameter is a useful tool which determines whether a time sequence of data presents long range dependency or not. Typically, the Hurst parameter values are between 0 and 1. If the Hurst parameter is above 0.5, then there is long range dependence between the frames. Further details on the Hurst parameter, its interpretation, and its calculation can be found in [12].

There are several methods to estimate the Hurst parameter. The most common methods are the R/S analysis, variance analysis, and periodogram analysis (see [13]). These estimators can be used for different levels of aggregation. We describe the computational details of each estimator.

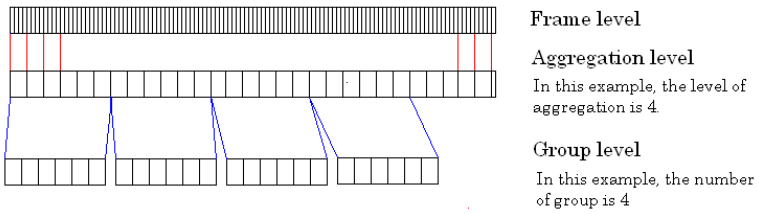


Figure 7: Steps of the R/S Analysis.

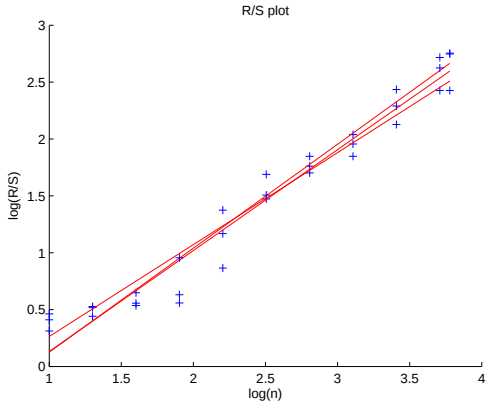


Figure 8: Combined R/S plots, for 3 different groups of a movie (Aggregation Level=1).

For the R/S estimator, we separate the video trace into groups (see Figure 7) and a R/S analysis is computed for each group (see equation 4).

$$W(X_n)_k = \sum_{i=1}^{k} X_n(i) - k * \bar{X_n}$$
$$\frac{R_X}{S_X}(n) = \frac{range(X_n)}{\sigma_{X_n}} = \frac{max(W(X_n), 0) - min(W(X_n), 0)}{\sigma_{X_n}} \tag{4}$$
$$\frac{R_X}{S_X}(n) \sim c_X * n^H$$

where X_n is the vector of the first n frames of the group. σ_{X_n} and $\bar{X_n}$ are respectively the standard deviation and the mean of X_n. H is the value of the Hurst parameter and c_X is a consant. '$\sim$' means that $\frac{R_X}{S_X}(n)$ is asymptotically porportional to $c_X * n^H$. Based on these values, we can plot the R/S plot with $\log(R/S)$ on the y-axis and $\log(n)$ on the x-axis. Using linear regression, we find the best fit to the data point (see Figure 8). Finally, the Hurst parameter is the average of the gradient found for each group. A tradeoff of this method is the choice of the values of n. For small n, short term correlations dominate and the readings are not valid. For large n, there are few samples and the value of $R/S(n)$ will not be accurate.

The variance estimator splits the video frame into blocks of length m. Then the variance is

computed (equation 5).

$$var(X^{(m)}) \sim c_X * m^{2H-2} \tag{5}$$

where $X^{(m)}$ is a time series derived from the original video trace (X) by aggregating it over blocks of size m. The sample variance $var(X^{(m)})$ is asymptotically proportional to $2H-2$ for large L/m and m, where L is the length of the video trace. We then plot $\log(var(X^{(m)}))$ versus $\log(m)$. Using linear regression, we find the best fit to the data point and we determine the gradient $(2H-2)$ of the plot. We deduce H from this value. A tradeoff of this method is the choice of the values of m.

In the periodogram method, we compute the periodogram of the video trace (see equation 6) and we plot $\log(I(\lambda))$ versus $\log(\lambda)$. Using linear regression, we find the best fit to the data points. The gradient of the straight line obtained is $1-2H$. We finally deduce H.

$$I(\lambda) = \frac{1}{2\pi L}\left|\sum_{j=1}^{L} X_j e^{ij\lambda}\right|^2 \sim c_X * \lambda^{1-2H} \tag{6}$$

where L is the length of the video trace and λ is the frequency.

The R/S analysis and the variance analysis are time-domain estimators. The periodogram analysis is frequency-domain estimator. Sometimes, these estimators are not accurate due to the non-stationarity of the scaling exponent. Moreover, the frequency-domain estimator is easily affected by the strong short-range dependency of the traces. These can lead to estimations of the Hurst parameter above 1 (see [14]).

In Table 3, we compare the results obtained with the three estimators. Note that the range of the values as well as the trends are the same for the three estimators. The periodogram estimator often exhibits values above 1 due to the strong short range dependence characteristic of the video traces. This estimator is not reliable in the case of video traces presenting strong short term correlation. The R/S estimator exhibits two values slightly above 1. They concern the Terminator 2 movie encoded with $QP = 10$ which presents very strong short term correlation (see section 3.4). An estimator of the long range dependence cannot be completely independent of the short term dependence. This leads us to consider the R/S and variance estimators as reliable estimator of the Hurst parameter.

Table 4 presents the estimation of the Hurst parameter using the R/S estimator for different movies encoded with G12/B2. These values are all above 0.5 and very often, above 0.7. This shows strong long range dependence between the frames. This characteristic can be exploited for the long term prediction.

A very important observation in Table 4 is that the Hurst parameter decreases when the quality of the video decreases, i.e. when the QP value increases. This shows that the long range characteristic

Table 3: Estimation of the Hurst Parameter with different methods for movies encoded with MPEG-4 AVC with G12/B2.

Movie	QP	Method	Level of Aggregation			
			1	**2**	**3**	**12**
Terminator 2	10	R/S Analysis	1.0173	1.0225	0.9480	0.8485
		Variance	0.8986	0.8444	0.7794	0.8065
		Periodogram	1.0540	1.1981	1.3600	1.4610
	28	R/S Analysis	0.9347	0.9783	0.9973	0.8424
		Variance	0.8983	0.8433	0.7533	0.6839
		Periodogram	0.8549	0.9963	1.1514	1.3875
	48	R/S Analysis	0.8650	0.9361	0.9730	0.8184
		Variance	0.8930	0.8301	0.6998	0.5948
		Periodogram	0.7319	0.8612	1.0025	1.3357
Horizon Talk Show	28	R/S Analysis	0.6921	0.7262	0.7603	0.7879
		Variance	0.8882	0.8145	0.7223	0.6938
		Periodogram	0.7194	0.8351	0.9611	1.1885
From Mars to China	28	R/S Analysis	0.8454	0.8679	0.8773	0.8244
		Variance	0.9104	0.8713	0.8390	0.8437
		Periodogram	0.7777	0.9189	1.0533	1.3351

is more important for the movies encoded in high quality than the ones encoded in low quality. We will see in Chapter 5 that the results for the long term prediction confirms this property.

Table 4: Estimation of Hurst Parameter with the R/S estimator for movies encoded with MPEG-4 AVC or MPEG-2 with G12/B2.

Movie	Standard	QP	Level of Aggregation						
			1	**2**	**3**	**12**	**24**	**48**	**96**
Terminator2	MPEG-4 AVC	10	1.0173	1.0225	0.9480	0.8485	0.8535	0.8093	0.8395
		28	0.9347	0.9783	0.9973	0.8424	0.8201	0.7535	0.7136
		48	0.8650	0.9361	0.9730	0.8184	0.7870	0.7971	0.7469
	MPEG-2	10	0.9705	1.0080	1.0259	0.8709	0.8310	0.7807	0.7593
		20	0.9321	0.9876	1.0094	0.8401	0.8024	0.7819	0.7777
		30	0.9040	0.9623	0.9973	0.8331	0.8045	0.7817	0.7755
Sony	MPEG-4 AVC	10	0.9772	1.0247	1.0247	0.9468	0.9311	0.8690	0.8171
		28	0.8864	0.9654	1.0102	0.9694	0.9809	0.9224	0.9064
		48	0.8081	0.8984	0.9884	0.9885	0.9615	0.9213	0.8857
	MPEG-2	10	0.9396	1.0266	1.1133	0.9691	0.9498	0.8882	0.8597
		15	0.8948	0.9839	1.0810	0.9713	0.9523	0.8909	0.8585
	Part2	8	0.8930	0.9862	1.0769	0.9690	0.9475	0.8838	0.8353
		12	0.8778	0.9729	1.0687	0.9783	0.9704	0.8949	0.8452
Horizon	MPEG-4 AVC	28	0.6921	0.7262	0.7603	0.7879	0.7563	0.7280	0.6952
Mars to China	MPEG-4 AVC	28	0.8454	0.8679	0.8773	0.8244	0.8180	0.8228	0.8240

CHAPTER 4

SHORT-TERM PREDICTION

The short term prediction can be used to allocate bandwidth in the short-term. In order to be used in a real time fashion, the models have to be very simple. The models proposed are based on the conclusions of the previous Chapter. We propose a model for the B-frame size prediction in section 4.1 and a model for the GOP size prediction in section 4.2. Finally we propose a algorithm using a scene change detector metric to enhance these models in section 4.3.

4.1 B-frame prediction

The B-frame size prediction is of particular interest due to the possibility of dropping the B-frames to conserve bandwidth [11].

4.1.1 Model

First, we introduce the following notations. $B_{1,t}$ is the size of the t^{th} B-1 frame (i.e., the size of the first B-frame of the t^{th} GOP of the encoded video). In the same way, we define the vectors B_2, B_3, B_4, B_5, B_6, B_7, B_8, P_1, P_2 and P_3 to correspond to the actual trace. All these vectors have the same length (L), which is the number of GOPs in the encoded video. The vector B is the trace of all the B-frames. $B_t = [B_{2,t}\ B_{3,t}\ B_{4,t}\ B_{5,t}\ B_{6,t}\ B_{7,t}\ B_{8,t}]$. The length of B is $8L$.

In the previous Chapter, we have shown that a B-frame is strongly correlated with the previous B-frame and the closest P-frame. Our model for the prediction a B-frame is based on these correlations. We employ a linear model (denoted as PB) for all t between 2 and L as described in equation 7. The model for the B-1 frames is slightly different because the B-frame used is not part of the same

GOP.

$$\begin{aligned}
\hat{B}_{1,t} &= \alpha_1 * B_{8,t-1} + \gamma_1 * P_{1,t} \\
\hat{B}_{2,t} &= \alpha_2 * B_{1,t} + \gamma_2 * P_{1,t} \\
\hat{B}_{3,t} &= \alpha_3 * B_{2,t} + \gamma_3 * P_{1,t} \\
\hat{B}_{4,t} &= \alpha_4 * B_{3,t} + \gamma_4 * P_{2,t} \\
\hat{B}_{5,t} &= \alpha_5 * B_{4,t} + \gamma_5 * P_{2,t} \\
\hat{B}_{6,t} &= \alpha_6 * B_{5,t} + \gamma_6 * P_{3,t} \\
\hat{B}_{7,t} &= \alpha_7 * B_{6,t} + \gamma_7 * P_{3,t} \\
\hat{B}_{8,t} &= \alpha_8 * B_{7,t} + \gamma_8 * P_{3,t}
\end{aligned} \tag{7}$$

Since the predicted value may not match the actual value exactly, we wish to quantify the accuracy of the model. For this purpose, we compute the marginal error of the prediction, defined as the difference between the predicted size of the frame and its actual size (equation 8).

$$\epsilon_t = \hat{X}_t - X_t \tag{8}$$

From this marginal error, we can define the Relative Percentage Error (RPE) as shown in equation 9.

$$RPE = \frac{\sum_{t=1}^{L} \epsilon_t}{\sum_{t=1}^{L} \hat{X}_t} * 100 \tag{9}$$

We employ a least squared method, implemented on Matlab, to find the best coefficients α_i and γ_i with i ranging from 1 and 8. After computing these coefficients, we use equation 7 to build the predicted vector for the B-frames.

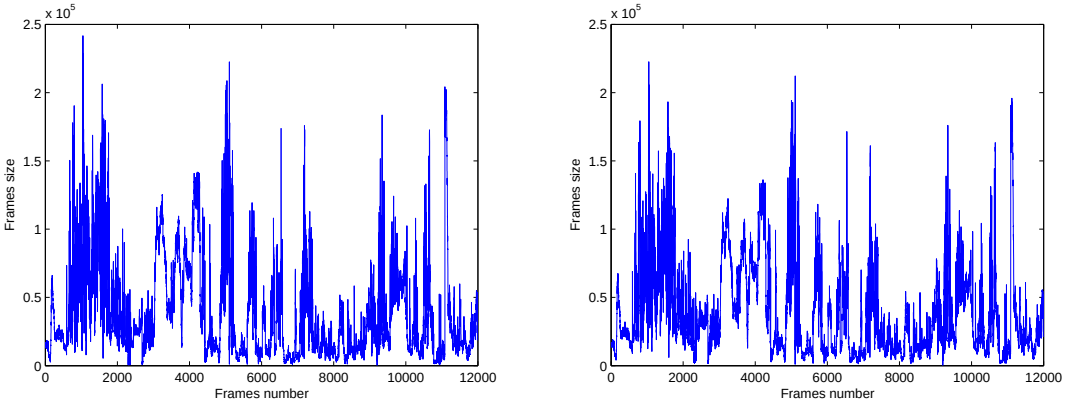
4.1.2 Results and comments

We perform our simulation on 4 different movies encoded in MPEG-4 AVC with a G12/B2 pattern and different values of QP. The Sony and the Terminator 2 movies are about 18,000 frames long. The Horizon Talk show and the From Mars to China movie are about 50,000 frames long.

The α_i and γ_i obtained with these movies are similar. Table 5 shows these coefficients for the Terminator 2 movie. Note that the values obtained for the B-1 and B-8 frames are quite different from the values obtained for the other B-frames. This is due to the fact that the model used is quite different. In this case, the P-frame used for the prediction of the B-1 and B-8 frames is two frames away from the B-frame of interest. As a consequence, the coefficients of correlation between these B- and P-frames are weaker - but are still the best.

Table 5: Values of α_i and γ_i obtained for the B-frame size prediction using PB model for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=28.

α	0.8436	0.8945	0.8568	0.9259	0.8878	0.881	0.8751	0.9311
γ	0.0579	0.0391	0.0429	0.027	0.0359	0.0375	0.0344	0.02



(a) Original Trace (b) Predicted Trace

Figure 9: B-frame trace for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=10.

Figures 9(a) and 9(b) shows the original and predicted trace for the movie Terminator 2 with QP=10. You can observe that the traces are very close to each other, which shows the accuracy of the model.

The statistics of the marginal error and the RPE for four movies are shown in Table 6. In most cases, the RPE is low, implying that the linear prediction model is accurate. However, the RPE for the Horizon Talk show movie is higher. This can be explained by the lower values for the

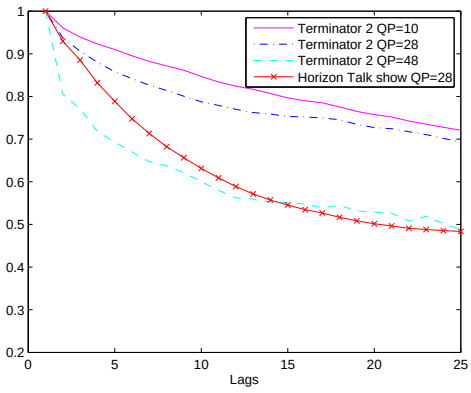


Figure 10: Autocorrelation of the B-frames for several movies encoded with MPEG-4 AVC with G12/B2.

Table 6: Marginal error statistics and RPE for the B-frame size prediction for movies encoded with MPEG-4 AVC with G12/B2.

	$\bar{\epsilon}$	$\bar{\lvert\epsilon\rvert}$	σ_ϵ	$\sigma_{\lvert\epsilon\rvert}$	**Max** ϵ	**Min** ϵ	**RPE**
			Terminator 2 QP=10				
PB	4.2E+04	-1.0E+03	6.3E+04	7.6E+04	1.3E+06	-1061100	5.43%
Ar	4.2E+04	-1.0E+03	6.4E+04	7.6E+04	1.3E+06	-1068700	5.45%
Ba	4.2E+04	-1.7E+03	6.4E+04	7.6E+04	1.3E+06	-1070200	5.44%
BB	3.9E+04	2.8E+03	6.7E+04	7.7E+04	1.4E+06	-1082400	5.14%
			Terminator 2 QP=28				
PB	5.7E+03	-9.8E+01	1.1E+04	1.2E+04	2.2E+05	-123840	14.79%
Ar	5.7E+03	-1.1E+02	1.1E+04	1.2E+04	2.2E+05	-125800	14.84%
Ba	5.7E+03	-1.2E+02	1.1E+04	1.2E+04	2.2E+05	-126250	14.78%
BB	5.4E+03	8.4E+02	1.1E+04	1.2E+04	2.2E+05	-140040	14.01%
			Terminator 2 QP=48				
PB	6.6E+02	-4.4E+01	1.4E+03	1.6E+03	2.7E+04	-15169	28.10%
Ar	6.7E+02	-4.9E+01	1.5E+03	1.6E+03	2.7E+04	-17884	28.63%
Ba	6.6E+02	-3.6E+01	1.5E+03	1.6E+03	2.7E+04	-18389	28.12%
BB	6.1E+02	1.7E+02	1.7E+02	1.5E+02	2.7E+04	-18994	26.10%
			Sony QP=48				
PB	2.30E+02	2.76E+01	4.49E+02	5.03E+02	17315	-9550	25.51%
Ar	2.42E+02	2.41E+01	4.95E+02	5.51E+02	17369	-14104	26.90%
Ba	2.33E+02	1.33E+00	5.08E+02	5.59E+02	17359	-15385	25.86%
BB	2.35E+02	6.63E+01	5.08E+02	5.55E+02	17377	-12963	26.04%
			From Mars to China QP=28				
PB	7.3E+03	-4.4E+02	1.5E+04	1.7E+04	5.5E+05	-411780	12.98%
Ar	7.3E+03	-4.4E+02	1.5E+04	1.7E+04	5.5E+05	-399230	12.96%
Ba	7.2E+03	-5.0E+02	1.5E+04	1.7E+04	5.5E+05	-403200	12.81%
BB	6.8E+03	1.2E+03	1.6E+04	1.7E+04	5.5E+05	-402270	12.14%
			Horizon Talk show QP=28				
PB	1.7E+03	-2.9E+01	3.1E+03	3.6E+03	1.2E+05	-75819	15.36%
Ar	1.7E+03	-2.0E+01	3.2E+03	3.6E+03	1.2E+05	-78873	15.39%
Ba	1.7E+03	-3.2E+01	3.2E+03	3.6E+03	1.2E+05	-78713	15.42%
BB	1.8E+03	2.9E+02	3.2E+03	3.6E+03	1.3E+05	-79790	15.79%

autocorrelation coefficients of the B-frames for the first lags, as shown in Figure 10. The higher RPE for the Terminator 2 (QP=48) movie is also explained by Figure 10. We observe that the autocorrelation of the B-frames for the movie with QP = 48 is weaker than the autocorrelation of the same movie encoded with QP=28 and QP=10. As a consequence, the linear prediction is less accurate. These results indicate that our proposed model performs best on movies encoded with low values of QP, which implies high quality movies.

It is also worth pointing out that the marginal error presents a long tail distribution, as shown in Figure 11. The algorithm presented in [2] can be used to fit a G-H distribution of equation 10 and the marginal distribution.

$$K_{g,h}(y) = A + C\left(\frac{e^{gy}-1}{g}\right)e^{h\frac{y^2}{2}} \tag{10}$$

where A and C are the location and scale parameters respectively, g and h are the skewness and

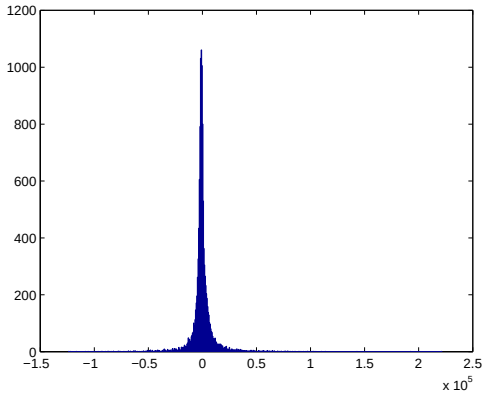


Figure 11: Distribution of the marginal error for the B-frames prediction for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=10.

kurtosis parameters, and y is a random variable with standard normal distribution.

4.1.3 Possible Enhancements

We can improve our model by simplifying its operation further. The least squared method provides 16 coefficients to describe the different B-frames. We would like to reduce this number to only one α and one γ to describe all B-frames. The α replaces each α_i and the γ replaces each γ_i in equation 7 .The values α and γ are determined using the values of α_i and γ_i determined previously. We propose two methods to compute them. In the arithmetic method, α and γ are simply the arithmetic average of the set of α_i and γ_i values. In the balanced method using equation 11, the values of α_i and γ_i are balanced by the inverse of the error they introduce into the model. We apply these two methods using the PB model for different movies and the results are presented in Table 6.

$$\begin{aligned} \alpha &= \sum_{i=1}^{8} \left(\frac{\alpha_i * \lambda_i}{\sum_i \lambda_i} \right) \\ \beta &= \sum_{i=1}^{8} \left(\frac{\beta_i * \lambda_i}{\sum_i \lambda_i} \right) \\ \lambda_i &= 1/\epsilon_i \end{aligned} \tag{11}$$

The results obtained with this method are nearly the same. The balanced method is only slightly more accurate than the arithmetic method but requires more processing. As such, the arithmetic method is preferred. Moreover, these results are very close to the values obtained without the arithmetic or balanced methods. As a consequence, the arithmetic method can be used to reduce the processing requirements of the predictive algorithm, which will in turn reduce the latency associated with its execution.

Finally, we have devised yet another model to predict the size of the B-frames using two previous

Table 7: Marginal error statistics and RPE for the GOP size prediction for movies encoded with MPEG-4 AVC with G12/B2.

	$\bar{\epsilon}$	$\bar{\lvert\epsilon\rvert}$	σ_ϵ	$\sigma_{\lvert\epsilon\rvert}$	**Max ϵ**	**Min ϵ**	**RPE**
				Terminator 2 QP=10			
IG	8.4E+05	8.7E+04	1.1E+06	1.4E+06	12245000	-8724900	7.37%
GG	8.4E+05	8.6E+04	1.1E+06	1.4E+06	12041000	-8594500	7.36%
				Terminator 2 QP=28			
IG	1.3E+05	1.4E+04	1.6E+05	2.1E+05	1977600	-1125800	14.71%
GG	1.3E+05	1.9E+04	1.6E+05	2.1E+05	2008600	-1159600	14.68%
				Terminator 2 QP=48			
IG	1.7E+04	1.9E+03	2.1E+04	2.7E+04	155460	-161790	16.67%
GG	1.7E+04	2.7E+03	2.1E+04	2.7E+04	155030	-164660	16.65%
				Sony QP=48			
IG	6.9E+03	7.4E+02	1.3E+04	1.5E+04	103910	-135030	7.49%
GG	6.9E+03	6.9E+02	1.2E+04	1.4E+04	105640	-122180	7.49%
				From Mars to China QP=28			
IG	2.3E+05	2.9E+04	3.7E+05	4.3E+05	5786200	-3951600	11.69%
GG	2.3E+05	3.9E+04	3.7E+05	4.4E+05	5756900	-3972800	11.69%
				Horizon Talk show QP=28			
IG	6.7E+04	3.1E+03	8.4E+04	1.1E+05	1908100	-750230	10.89%
GG	6.9E+04	8.3E+03	8.9E+04	1.1E+05	1806600	-1021500	11.19%

B-frames (denoted BB). The model is reasonable given the strong values of the autocorrelation of the B-frames for the first lags. The results are shown in Table 6. As can be seen, the BB method performs slightly better than the PB model more often.

4.2 GOP prediction

We have also constructed a model to predict the size of the GOP in the short term. Similar to the process for the B-frame size prediction, we use the conclusions of Chapter 3 to construct the model. We implemented two models. The first model (denoted GG) is described by equation 12 and uses the size of the two previous GOP as predictors. The second model (denoted IG) is described by equation 13 and uses the size of the previous GOP and the size of the I-frame of the previous GOP as predictors. The result are presented in Table 7.

$$\hat{G}_t = \alpha * G_{t-1} + \gamma * G_{t-2} \tag{12}$$

$$\hat{G}_t = \alpha * G_{t-1} + \gamma * I_{t-1} \tag{13}$$

where G_t is the size of the t^{th} GOP and I_t the size of the t^{th} I-frame. The coefficients α and γ are computed for all GOPs of the whole movie using a least squared method.

The results in Table 7 prove that linear prediction is accurate for predicting the GOP's sizes. As for the B-frames prediction, the model using the autocorrelation is more accurate than the model

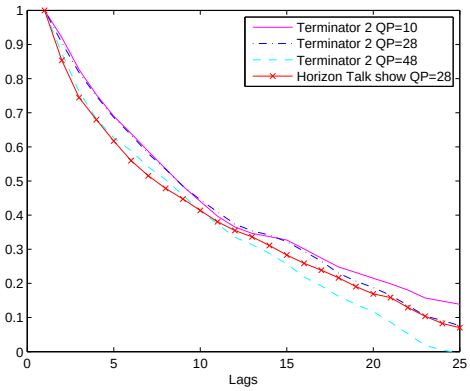


Figure 12: Autocorrelation of the GOP for several movies encoded with MPEG-4 AVC with G12/B2.

using the crosscorrelation. The I-frames have, in the GOP prediction, have the same role as the P-frames in the B-frame prediction.

Moreover, we can make the same conclusions for the high RPE for the Terminator 2 movie with QP=48. In Figure 12, the autocorrelation for the first lags of the movie QP=48 is noticeably weaker than for the same movie with QP=10 and QP=28, which explains the higher RPE.

This GOP prediction can be useful in allocating the required bandwidth in the longer term than allocating the bandwidth for the next B-frame.

4.3 Scene Change Detector

In this section we propose to use a metric to enhance the previous models. In these previous models, in order to find the best α and γ, we employ the least squared method for all the frames. However, during a movie, there may be scene changes, which typically lead to a significant change in the size of the frames. This may impair the accuracy of the prediction method. Previous work on scene changes has been reported (e.g., [6]). However, the majority of these methods are proposed for pre-recorded videos. In this section, we propose a new real-time algorithm to deal with scene changes. First, we use a simple metric to detect a scene change. Then, each time a scene change occurs, our algorithm will "reset" and treat the frames like a new movie that was transmitted until the next scene change. The video trace is partitioned according to the scene changes. We are going to apply this algorithm for the B-frame and GOP size prediction. The models are the ones presented in sections 4.1 (BB model) and 4.2 (GG model) respectively.

In order for our algorithm to be executed in real time, it is necessary to develop a simple scene change algorithm. The algorithm we propose is shown in Figure 13 (implemented in Matlab). In

```
nb_change=0;
place_change=[1];
for j=2:nb_GOP
    diff(j)=abs( I(j) - I(j-1) );
    if diff(j) > Threshold
        if ( j - place_change(end) > 3 )
            place_change = [place_change i];
            nb_change = nb_change + 1;
        end
    end
end
```

Figure 13: Scene Change Detection Algorithm.

the algorithm, *nb_change* is the number of detected scene changes in the movie; *place_change* is a vector which contains the frame numbers that indicate the occurrence of the scene changes; and *I(j)* is the size of the j^{th} I-frame.

This new algorithm computes the difference in the size of two adjacent I-frames. Each time the encoder receives an I-frame, it computes the difference in size with the previous I-frame it has received. If the difference is higher than a pre-defined threshold, the algorithm will start over. This metric is relevant because each scene change typically occurs at the beginning of a GOP. Moreover, the I-frames are computed without using any reference frame. As a consequence, their size reflects most of the scene changes. A key consideration is to choose a good threshold. If the threshold is too high, the impact of this method may not be significant because the encoder will only detect a few scene changes, although these changes tend to be abrupt scene changes. On another hand, if the threshold is too low, the algorithm will have to start over very often, and this can prevent the real-time transmission of the video. Therefore, the choice of the threshold is a tradeoff between the efficiency of the algorithm and real-time transmission. Note that even though a lower threshold enables the detection of gradual scene changes (in addition to abrupt changes), and hence better performance, this may not be necessary since our linear prediction algorithm can cope better with gradual scene changes than abrupt changes. In order to implement the linear prediction method, we have to ensure that every segment of the movies is at least 3 GOPs in length.

We apply this algorithm to several movies using several threshold values. The results are presented in Table 8. Clearly, the RPE of all movies encoded in all qualities decreases when the scene change detector algorithm is applied. As expected, better results are always obtained when the threshold is low. However, this may incur more latency due to the need to restart the algorithm regularly.

Table 8: RPE for the B-frame and the GOP size prediction for movies encoded with MPEG-4 AVC with G12/B2 when the scene change model is applied with different thresholds.

Threshold	Number of scene changes	Mean number of GOPs per segment	B-frames RPE	GOP RPE
Terminator 2 QP=10				
4000	367	5	2.06%	0.12%
10000	359	5	2.11%	0.21%
20000	342	5	2.18%	0.52%
50000	287	6	2.48%	1.34%
Terminator 2 QP=28				
4000	337	5	6.24%	1.21%
10000	288	6	6.82%	2.65%
20000	227	7	7.52%	4.34%
50000	125	13	9.84%	8.12%
Terminator 2 QP=48				
4000	206	8	14.13%	6.17%
10000	115	14	18.78%	10.22%
20000	44	33	22.30%	13.79%
50000	5	206	27.87%	16.26%
Sony QP=48				
4000	83	18	14.24%	4.30%
10000	41	35	16.93%	5.35%
20000	18	77	19.07%	6.56%
50000	4	302	20.48%	6.85%
From Mars to China QP=28				
4000	990	5	4.88%	0.86%
10000	894	5	5.25%	2.05%
20000	774	6	5.82%	3.18%
50000	568	8	7.06%	4.94%
Horizon Talk show QP=28				
4000	822	5	7.66%	2.64%
10000	622	7	9.20%	4.81%
20000	391	11	11.00%	7.07%
50000	129	32	13.53%	9.41%

We now compare the RPE obtained with this scene change detector applied to the B-frames prediction with the ones presented in Table 6. For the Terminator 2 (QP=10) movie, the model presented in section 3 yields a 5.43% RPE. Using the scene change detector, an RPE of 2.48% is obtained with a 50,000-byte threshold. The lowest RPE of 2.06% is achieved with a threshold of 4,000 bytes. In this example, the choice of the threshold has almost no influence on the RPE. For the Terminator 2 (QP=48) movie, a 50,000-byte threshold is almost useless (the RPE decreases from 28.10% to 27.87% with the scene change detector algorithm). However, if we use a 4000-bytes threshold for the same movie, the RPE decreases significantly from 28.10% to 14.13%. The improvement obtained by the use of the scene change detector is also obvious for the GOP prediction. We can observe this by comparing the results in Table 7 and 8.

CHAPTER 5

LONG-TERM PREDICTION

The long range dependence characteristic of MPEG-4 AVC encoded videos has been shown in Chapter 3. We wish to exploit this property to predict the size of the frames in the long-term. In the long term, there may be many scenes changes between prior data used for prediction and the data we wish to predict. As a consequence the notions of GOP and crosscorrelation is not as important as in the short term. The different frame types (i.e. I-, P-, B-frames and GOP) are treated separately. This means that the I-frame sizes are predicted using the size of previous I-frames, the P-frame sizes are predicted using the size of previous P-frames, the B-frame size are predicted using the size of previous B-frames, and the GOP size are predicted using the size of previous GOPs. The algorithm for each frame type is the same, except that it does not use the same data as predictors.

Clearly, a long-term prediction algorithm may suffer degraded prediction accuracy and the higher complexity may result in higher latency. However, this is offset by the additional time available for long-term prediction and the need to forecast bandwidth usage well ahead of time in order to minimize packet losses during periods of peak bandwidth demands.

This study has been made for a single video stream. When the streams of many videos are multiplexed, the long-range dependence characteristic tends to decrease and the prediction becomes less accurate. First, we present the model in section 5.1. Then we provide the solution to the problem in section 5.2. Finally, the influence of M and δ on the results are discussed in section 5.3.

5.1 Presentation of the Model

We wish to use M video frames to predict the size of a video frame placed δ frames later. The model is shown in Figure 14. We wish to predict the video frame in black using the M video frames in grey. This group of frames is separated from the single frame to be predicted by δ frames. M determines the complexity of the model. A bigger M, results in a more complex model. On the other hand, if M is too small, the accuracy of the prediction may be compromised. δ defines the notion of "long-term" prediction. If δ is small, the prediction corresponds to short-term prediction. A bigger δ leads to long-term prediction. However, even if the Hurst parameter shows strong long range dependence, the correlation between the frames decreases when the lag increases, as illustrated in Figure 6. As reasonable value δ is less than 1200 frames.

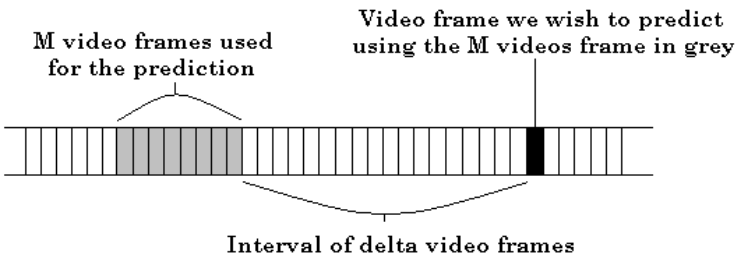


Figure 14: Model of the Long-term Prediction.

5.2 Equations and solution of the Model

We chose a linear model as presented in equation 14.

$$\hat{x}(t+\delta) = \sum_{i=1}^{M} g(i) * x(t+i-M-1) \tag{14}$$

In this equation $x(t+\delta)$ corresponds to the video frame in black in Figure 14. $x(t-M)$ to $x(t-1)$ corresponds to the video frames in grey. This equation is only valid for values of t which satisfy $1 < t+\delta < L$ and $1 < t-M < L$ at the same time. The inequalities can be combined as: $1+M < t < L-\delta$. L is the length of the data considered. So L correponds to the number of GOPs of the video if we consider the I-frames or the GOP, L is three times the number of GOPs if we consider the P-frames, and L is 8 times the number of GOPs if we consider the B-frames.

Now, we need to define the error E and minimize this error, in order to determine the coefficients $g(i)$. E is defined in equation 15

$$E_{\delta,M} = \sum_{t=1+M}^{L-\delta} \epsilon_{\delta,M}^2(t+\delta) \qquad \text{with} \qquad \epsilon_{\delta,M}(t+\delta) = x(t+\delta) - \hat{x}(t+\delta) \tag{15}$$

ϵ is the same marginal error as defined for the short-term prediction.

In order to minimize E, we set its derivative to zero, as shown in equation 16.

$$\forall i = 1,...,M : \frac{\partial E_{\delta,M}}{\partial g(i)} = 0 \Leftrightarrow \sum_{t=1+M}^{L-\delta} \epsilon_{\delta,M}(t+\delta) * \frac{\partial \epsilon_{\delta,M}(t+\delta)}{\partial g(i)} = 0$$

$$\Leftrightarrow \sum_{j=1}^{M} g(j) * r_{xx}(i,j) = r_x(i)$$

where the matrix R_{xx} and the vector R_x are defined as follow: (16)

$$\forall i,j = 1,...,M : r_{xx}(i,j) = \sum_{t=j}^{L-\delta+j-M-1} x(t) * x(t+i-j)$$

$$\forall i = 1,...,M : r_x(i) = \sum_{t=1+M+\delta}^{L} x(t) * x(t+i-M-\delta-1)$$

Finally, the coefficients are determined by:

$$[g(1)\ g(2)\ ...\ g(M)]^T = R_{xx}^{-1} * R_x. \tag{17}$$

As for the short-term prediction, once the optimal coefficients are known, we use the equation 14 of the model to construct the predicted vector.

Note that if we set $\delta = 1$ and $M = 2$, we have the BB or GG models presented in Chapter 4, and the computation of the coefficients $g(1)$ and $g(2)$ is equivalent to the least squared method used for the computation of the coefficients α and γ.

5.3 Study of the parameters M and δ

The influence of M and δ are detailed in this section.

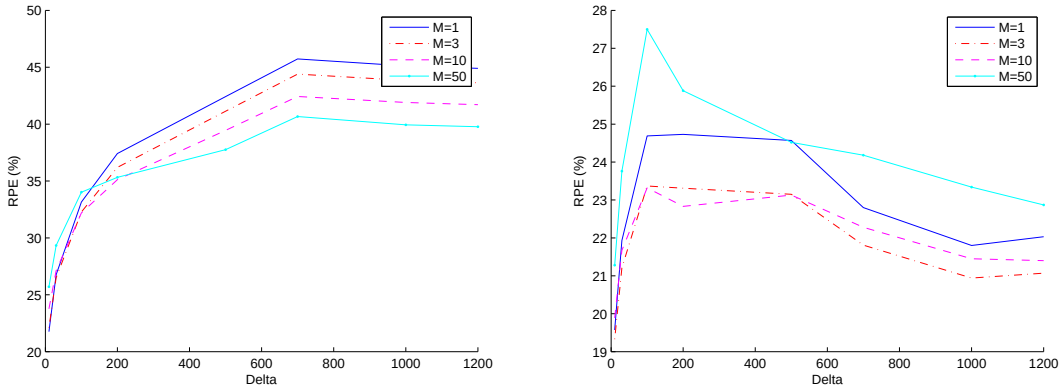
5.3.1 Influence and choice of the parameter M

First of all, we observe the influence of the parameter M. We have implemented the algorithm with values of $M = 1, 3, 10, 50$. The plot of the RPE function of δ is shown in Figure 15. For each frame type, the algorithm performs the best with $M = 1$ or with $M = 3$. Moreover, $M = 3$ and $M = 1$ reduces the complexity of the algorithm. The choice of $M = 3$ performs slightly better. Therefore, we chose $M = 3$ for the remaining studies. Other plots depicting the influence of M can be found in Figure 16 and Figure 17.

5.3.2 Influence of the parameter δ

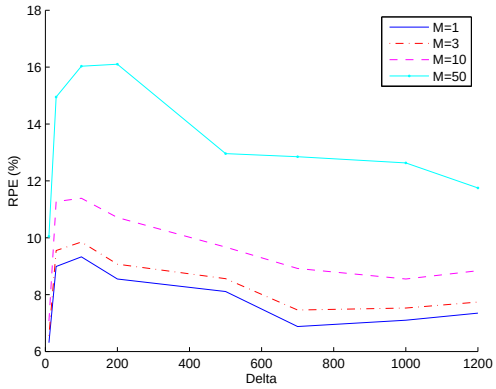
The parameter δ is more delicate and at the same time more interesting to study. In Figure 18, the plots of the RPE function of δ are presented for 5 different movies. For the I-frame and GOP size prediction, the RPEs first increase with δ and then decrease when δ is above 200 frames. For the P-frames, the RPE is quite constant when δ is above 200 frames. For the B-frames, the RPE always increases when δ increases. The B-frames differ from the other frames and the GOP in two ways. First, there are more B-frames than other frame types, and secondly, the B-frames contain only temporal information, which are more subject to change during a movie than spatial information. The P-frames are intermediate between the B-frames, and the I-frames and the GOP. This explains the difference in the RPE for the different frame types when δ increases.

Note that the long-term prediction for the B-frame of the movie Terminator 2 QP=10 is very good. The RPE is close to 10% for every value of δ. For the short-term prediction the RPE for that movie is close to 5%.

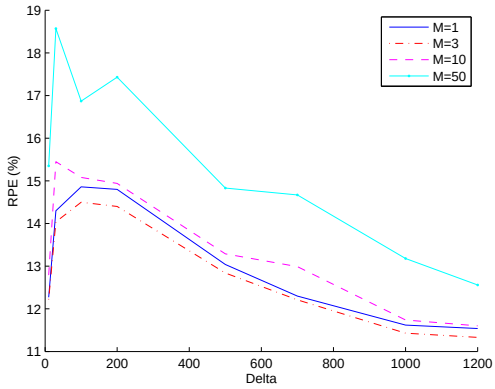


(a) B-frames

(b) P-frames



(c) I-frames

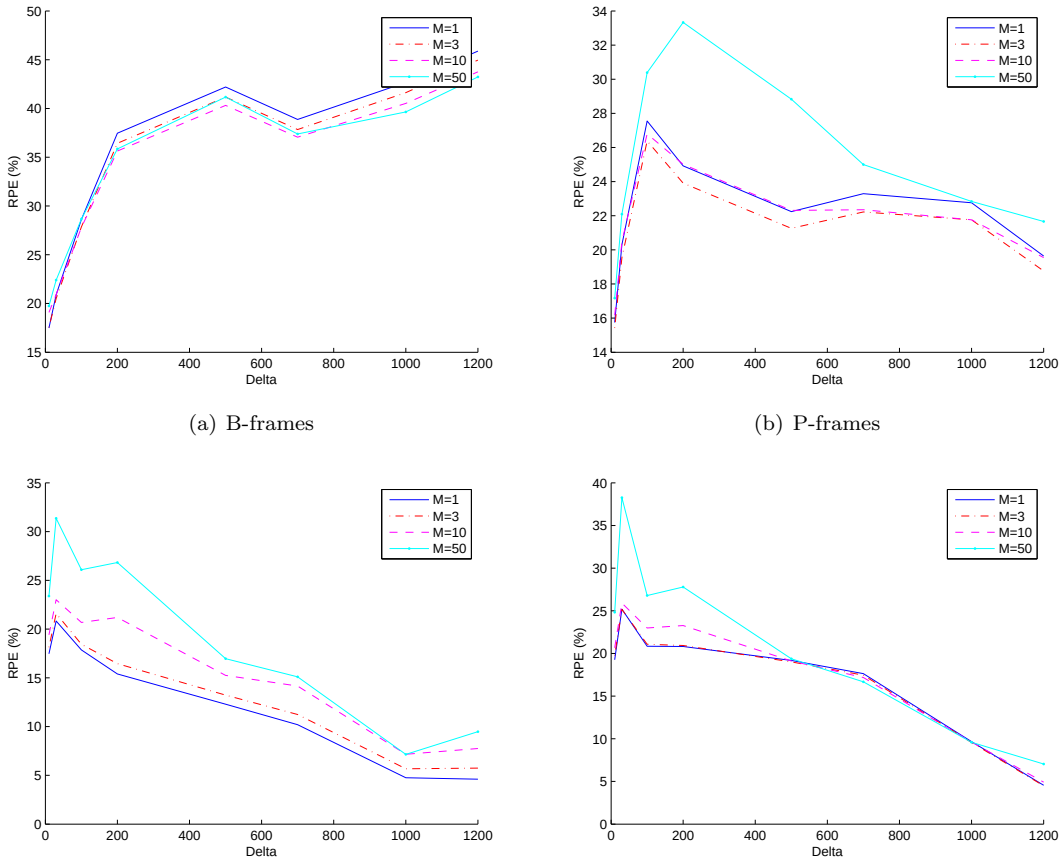


(d) GOP

Figure 15: Influence of M on the RPE of the long term prediction for Horizon Talk show movie encoded with MPEG-4 AVC with G12/B2 and QP=28.

In general, the GOP size prediction is better than the I-frame size prediction, which is better than the P-frame size prediction. This is consistent with the values of the CoV obtained in Chapter 3 for the three frame types. The CoV is higher for the B-frame than the P-frame, for the P-frame than the I-frame, and for the I-frames than the GOP.

To summarize, the long-term prediction model is accurate for the P- and I-frames, and the GOP in most cases. For the B-frame prediction, the model is accurate if the CoV is low and the Hurst parameter is high. This is typically the case for the high quality movies (i.e. movies encoded with a low QP).



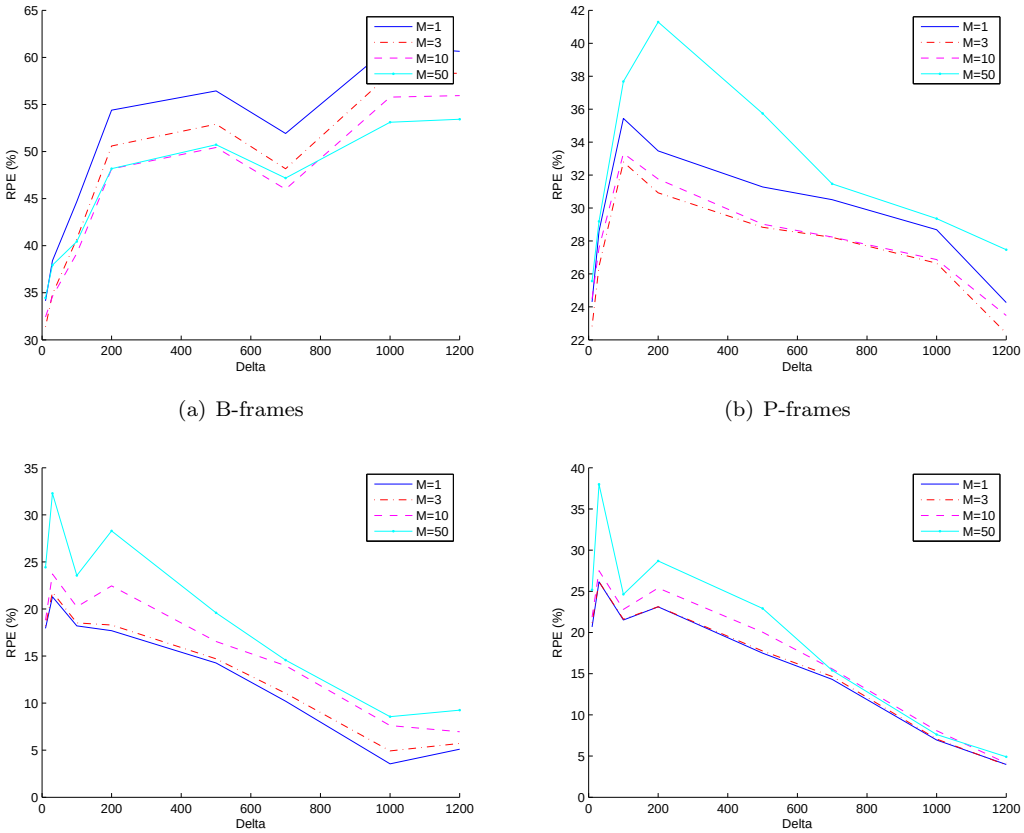
(a) B-frames

(b) P-frames

(c) I-frames

(d) GOP

Figure 16: Influence of M on the RPE for the long term prediction for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=28.



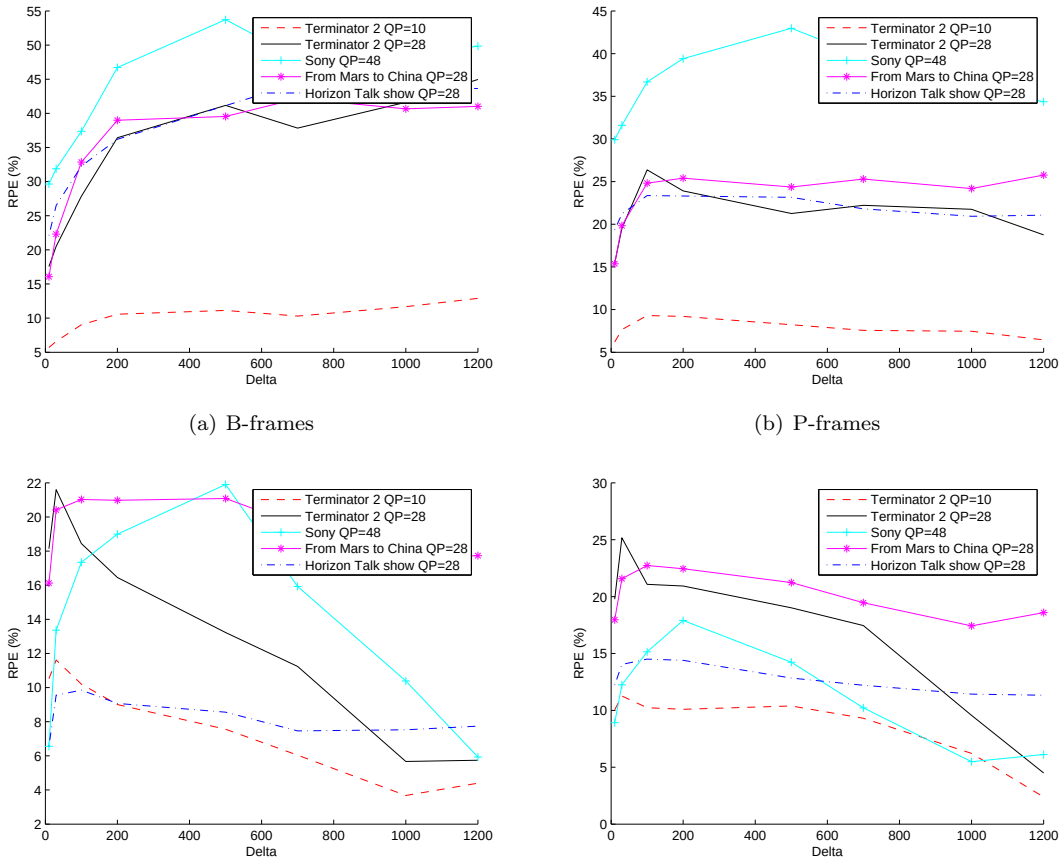
(a) B-frames

(b) P-frames

(c) I-frames

(d) GOP

Figure 17: Influence of M on the RPE for the long term prediction for Terminator 2 movie encoded with MPEG-4 AVC with G12/B2 and QP=48.



(a) B-frames

(b) P-frames

(c) I-frames

(d) GOP

Figure 18: Influence of δ on the RPE for the long term prediction for several movies encoded with MPEG-4 AVC with G12/B2.

CHAPTER 6

IMPACT OF VIDEO QUALITY AND VIDEO STANDARD ON LONG-TERM PREDICTION ACCURACY

Currently, MPEG-2 is widely used in Broadband Cable Networks. However, many cable companies are now migrating to MPEG-4 AVC. This chapter compares the results of the long-term prediction for the two standards and takes into account the quality of the movie transmitted.

The video traces of the two standards and different QP values can be obtained from [8]. We focus on the Terminator 2 movie, which provided excellent prediction results in Chapters 4 and 5. The plots of the RPE are presented on Figure 19. The values of the QP for the MPEG-4 AVC standard are QP=10, 22, 28, 34 and 48. The values of the QP for the MPEG-2 standards are QP=10, 15, 20, 25 and 30.

We first analyze the influence of QP for each standard in section 6.1. Then we compare the two standards for movies encoded with the same QP (section 6.2) and for movies which have the same size (section 6.3). Finally a global comparison of MPEG-4 AVC and MPEG-2 is made in section 6.4

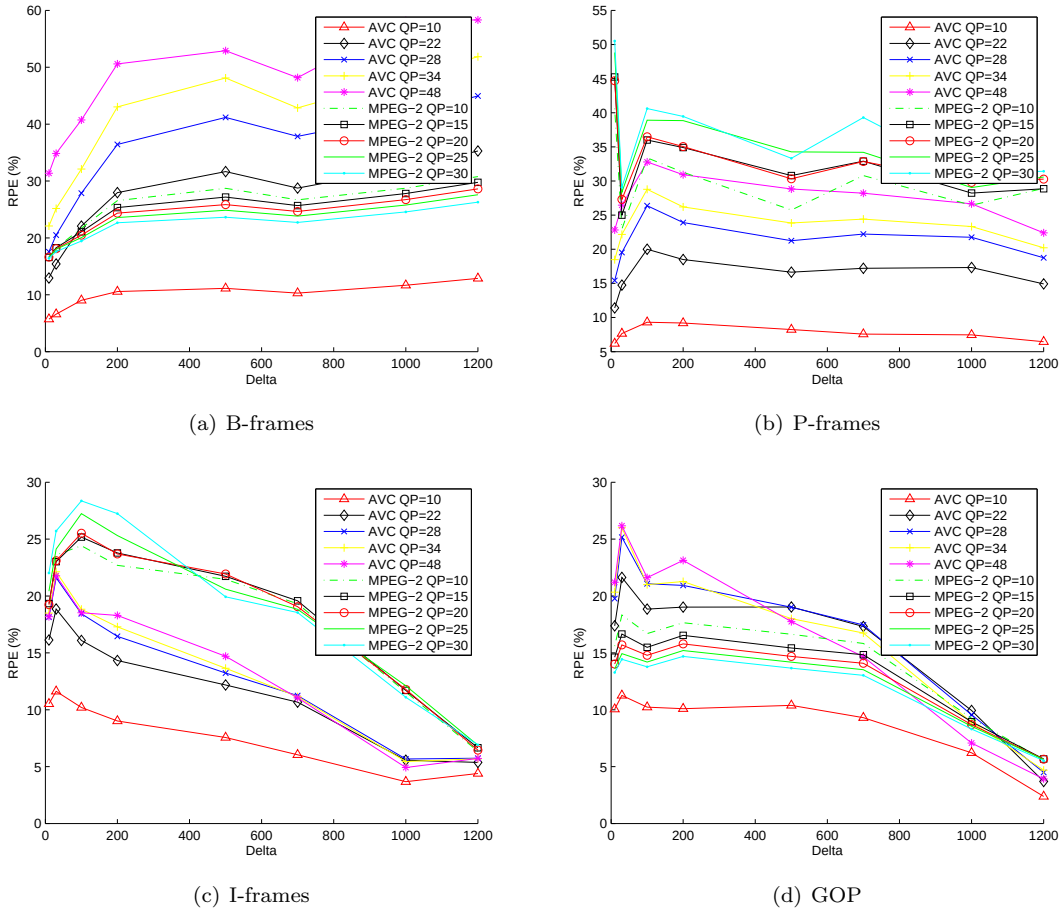
6.1 Impact of the QP value for each standard

The accuracy of the prediction for videos encoded with MPEG-4 AVC increases when the QP decreases, that is when the quality increases. This is consistent with the values of the Hurst parameter obtained in Table 9. The Hurst parameter may decrease when the quality decreases. This means that the long range characteristic of the video degrades when the video is encoded with a high QP. As a consequence, the long term prediction is less accurate when the QP increases.

However, this is not the case for videos encoded with MPEG-2. The Hurst parameter decreases when the QP value increases for these videos. For MPEG-2 video, the accuracy of the prediction increases with the quality of the video for the P- and I-frames. For the GOP and the B-frame size prediction, the accuracy of the prediction decreases with the quality of the video.

6.2 Comparison of MPEG-2 and MPEG-4 AVC movies with same QP value

Two videos with the same QP present approximately the same quality. When we compare videos encoded with a small QP (QP=10), the prediction of the data of the MPEG-4 AVC video is more



(a) B-frames

(b) P-frames

(c) I-frames

(d) GOP

Figure 19: Influence of the qualities and the standard of a video on the RPE for the long term prediction for Terminator 2 encoded with MPEG-4 AVC and MPEG-2 with G12/B2.

accurate in all cases (I-, P-, B-frame and GOP size prediction). However, if we compare videos encoded with higher QP than 10, we observe that the prediction is more accurate for MPEG-4 AVC videos for the P- and I-frames and for MPEG-2 videos for the B-frames and the GOP. We observe that when comparing MPEG-4 AVC with QP=22 and MPEG-2 with QP=20 and 25, as well as MPEG-4 AVC with QP=28 and MPEG-2 with QP=25 and 30.

6.3 Comparison of MPEG-2 and MPEG-4 AVC movies with same size

Videos with approximately the same total size use approximately the same amount of bandwidth when they are transmitted. The total size for the two standards and the different QP values for Terminator 2 are presented in Table 9. We compare the video encoded with MPEG-4 AVC with QP=28 to the video encoded with MPEG-2 with QP=20. Their sizes are both approximately $1.6*10^8$ bytes. We observe that the B-frame and the GOP size prediction is more accurate for the video encoded with MPEG-2 with QP=20, and that the I- and P-frame size prediction is more accurate

Table 9: Terminator 2 Statistics.

Standard	QP	Sum of Frame Sizes (Bytes)	Hurst Parameter Level of agg: 1
H.264	**10**	2.15E+09	1.0173
	22	3.81E+08	0.9682
	28	1.66E+08	0.9347
	34	8.17E+07	0.9032
	48	1.89E+07	0.8650
MPEG-2	**10**	2.42E+08	0.9705
	15	1.87E+08	0.9509
	20	1.59E+08	0.9321
	25	1.43E+08	0.9157
	30	1.32E+08	0.9040

for the video encoded with MPEG-4 AVC with QP=28.

6.4 Global comparison of MPEG-2 and MPEG-4 AVC

The prediction of the size of the I- and P-frames works better when the video is encoded with MPEG-2 than with MPEG-4 AVC. However, the prediction of the size of the B-frames and the GOP works better when the video is encoded with MPEG-4 AVC than with MPEG-2. The conclusion about these comparisons is that the video encoded with MPEG-4 AVC with QP=10 gives the best performance for the I-, P-, B-frame and GOP prediction. With a good choice of δ (between 50 and 1000 frames), the RPE is below 11%.

CHAPTER 7

FUTURE WORK: VIDEO NETWORK CODING

Network coding has been proposed eight years ago by a small group of researchers [15]. Since then, it has attracted much interest. It has the potential to improve the efficiency of information transmission in packet networks. It promises to have a significant impact on both the theory and practice of network design.

The key idea of network coding is to allow mixing/encoding of data at intermediate network nodes. It generalizes the operation of intermediate nodes in the network. They are allowed to not only forward but also combine the incoming independent information flows. Basically, it changes their operation from routing to coding.

With source coding, network routers typically treat packets individually, as a piece of a message. Each packet is switched along appropriate network pathways. The receiver reassembles them into a complete message. With network coding, the network coders mix/encode these packets at intermediate nodes. The receiver gathers them and when a sufficient number of mixed packets has been received, it can infer which message has been sent.

With network coding, routers can mix packets from different flows in a way that it increases the information content of each transmission and therefore, the throughput. Network coding reduces delay and improves the robustness of the transmission. It can also provide benefits in terms of complexity, scalability, and security. It can be used to realize energy saving [16]. Network coding can be applied in different traffic configurations: multicasting or multiple unicasts.

This field has recently found commercial applications in content distribution, peer-to-peer design, and enabling high-throughput wireless networks [17]. It can address the problem of bandwidth limitation. In wireless networks the coding enables the routers to compress the information whenever possible to reduce the number of transmissions required to deliver the packets in the router's queue. A fewer number of transmissions translates directly to less bandwidth consumption and higher throughput.

The authors in [18] apply network coding to video transmission. They present new schemes to be applied to near-Video-on-Demand (nVoD) systems to address the problem of a lossy network. It proposes to use the redundancy of nVOD protocols to provide implicit error correction. The novel

approach combines segments of the original content without any overhead. Previously downloaded segments allow recovering from future packet losses. It is named implicit error correction to contrast with the explicit error correction scheme which adds an overhead to every segment transmitted. Multicast channels do not transmit original segments nor independently encoded ones. Instead, the protocol combines original segments as macro-blocks to produce encoded segments with zero overhead. This scheme cannot recover packets beyond a certain packet loss probability p, but that probability increases whenever the client obtains a new segment. Using this implicit error correction, the transmission is more robust, without requiring more bandwidth.

To conclude, network coding is new field of research which promises to attract a lot of interest in future, and can be applied to video transmission.

CHAPTER 8

CONCLUSION

In this thesis, we have analyzed the characteristics of MPEG-4 AVC video traffic in the short term and in the long term. We have presented a model for predicting, in the short-term, the size of the B-frame and the GOP of MPEG-4 AVC compressed video. It has been shown the model provides highly accurate prediction, in particular, for movies encoded with high quality resolution. Although the prediction models are very simple to implement, it can be simplified further using the balanced or arithmetic methods without sacrificing prediction accuracy. We also improved the accuracy of the linear prediction model using a scene change detector. We have shown that the relative percentage error can be lowered for any type of movie using this enhancement.

We have also presented a model for predicting the size of all frame types (I-, P-, B-frame and GOP) in the long-term. The accuracy of the prediction is excellent for movies encoded with MPEG-4 AVC with a small QP value. We have applied the model to MPEG-2 movies as well. It has been shown that in most cases, the prediction works better for the MPEG-4 AVC movies than for the MPEG-2 movies. This provides an incentive for using the new video standard to transport video with high quality resolution.

REFERENCES

[1] G. Van der Auwera, P. T. David, M. Reisslein, "Traffic Characteristics of H.264/AVC Variable Bit Rate Video", *Technical Report, Arizona State University*, September 2007. http://www.fulton.asu.edu/m̃re/COMMAG_06_00562_revised.pdf (october 2007).

[2] Z. Liu, et. al, "Periodic Data Traffic Modeling and Prediction-Based Bandwidth Allocation", *IEEE/ACM CNSR Conference*, April 2006.

[3] A. Lombardo, G. Morabito, G. Schembra, "An Accurate and Treatable Markov Model of MPEG-Video Traffic", *IEEE INFOCOM*, 1998, pp. 217-224.

[4] Y. Dong, Z. Zhang, D. Du, "Maximizing the Profit of VOD Service on Broadband Cable Networks", *IEEE Globecom*, December 2003.

[5] W. Xu and A.G. Qureshi, "Adaptive Linear Prediction of MPEG Video Traffic", *IEEE ISSPA* 1999, pp. 67 - 70.

[6] S. Feng and R. Sankar, "Limitation and Improvement to Linear Prediction and Smoothing-based Bandwidth Allocation for VBR Traffic", *IEEE Globecom* 1999, pp. 209 - 213.

[7] J. Beran, R. Sherman, M. Taqqu, W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic", *IEEE Transactions on Communications*, April 1995.

[8] Video Trace Resource, http://trace.eas.asu.edu/hd/index.html (january 2008).

[9] G. Davidson, et. al, "ATSC Video and Audio Coding", *Proceedings of IEEE*, Vol. 94, No. 1, January 2006, pp. 60 - 76.

[10] B. Bing and L. Lanfranchi, "Optimizing Video Transmission for Broadband Cable Networks", *IEEE CCNC*, January 10 - 12, 2008.

[11] C. Bhadricha, B. Bing, "Scalable MPEG-4 Video Transmission using Video Quality Metrics", *IEEE/ACM CNSR Conference*, May 2008.

[12] J.C. Cano, Pietro Manzoni, "On the use and calculation of the Hurst Parameter with MPEG video data traffic", *Euromicro Conference*, 2000.

[13] R. Clegg, "A practical guide to measuring the Hurst Parameter" *International Journal of Simulation: Systems, Science & Technology*, Octobre 2006.

[14] N. Cackov, Z. Lučić, M. Bogdanov, L. Trajković, "Wawelet-Based Estimation of Long-Range Dependence in MPEG Video Traces", *IEEE ISCAS*, May 2005.

[15] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network Information Flow", *IEEE Transactions on Information Theory*, IT-46, pp. 1204-1216, 2000.

[16] C. Fragouli, J. Widmer, and J.Y. Le Boudec, "Efficient Broadcasting Using Network Coding", *IEEE/ACM Transactions on Networking*, Vol. 16, No. 2, April 2008.

[17] D. Katabi, S. Katti, W. Hu, H. Rahul, and M. Medard, "On Practical Network Coding for Wireless Environments", *International Zurich Seminar on Communications*, February 2006.

[18] R. Asorey-Cacheda, F.J. González-Castaño, "Zero-Overhead Implicit Error Correction for nVoD", *IEEE CCNC*, January 10 - 12, 2008.

VITA

Laetitia I. Lanfranchi was born in Nice, France, in 1986. She completed her Diplôme d'Ingénieur Degree, from Supélec, Paris, France. She is currently working towards a Masters degree (with Thesis option) in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. Her research interests include video transport optimization, bandwidth prediction, and cable networks.

List of Publications:

- B. Bing and L. Lanfranchi, "Optimizing Video Transmission for Broadband Cable Networks", *IEEE CCNC*, January 10 - 12, 2008.

- L. Lanfranchi and B. Bing, "Short-Term MPEG-4 AVC Bandwidth Prediction for Broadband Cable Networks", *IEEE/ACM CNSR Conference*, May 2008.

- L. Lanfranchi and B. Bing, "High-Definition MPEG-4 AVC Traffic Analysis and Bandwidth Prediction", *IEEE ICCCN Conference*, August 2008.

- L. Lanfranchi and B. Bing, "MPEG-4 Bandwidth Prediction for Broadband Cable Networks", *IEEE Transactions on Broadcasting*, Fall 2008.